

CLAIMS

What is claimed is:

1. An adaptive method for obtaining representative text items from a plurality of text items in an active task, each of the plurality of text items having a plurality of attributes, the method comprising the steps of:

(a) for each of the plurality of text items, identifying each of the plurality of attributes;

(b) for each of the plurality of text items, assigning a weight to each of the plurality of attributes, wherein at least two of the plurality of attributes are assigned different weights;

(c) for each of the plurality of text items, calculating an accumulated weight by accumulating the weight assigned to each of the plurality of attributes; and

(d) ranking the plurality of text items based on the accumulated weight of each of the plurality of text items.

2. The method of claim 1, wherein the plurality of attributes includes at least one of style, font size, and location of the text item.

3. The method of claim 2, wherein a weight assigned to a word is increased in response to the word locating in a specific region.

4. The method of claim 3, wherein the specific region is an active window being viewed by a user or a region selected by the user.

5. The method of claim 1 further forming a plurality of search terms based on a result of the ranking step.

6. A data processing system for obtaining representative text items from a plurality of text items in an active task, each of the plurality of text items having a plurality of attributes, the system comprising:

a data processor for processing data;

a data storage device for storing instructions; and

a data transmission path coupled to the data processor and the data storage device;

wherein the instructions, when executed by the data processor, controls the data processing system to perform the machine-implemented steps of:

(a) for each of the plurality of text items, identifying each of the plurality of attributes;

(b) for each of the plurality of text items, assigning a weight to each of the plurality of attributes, wherein at least two of the plurality of attributes are assigned different weights;

(c) for each of the plurality of text items, calculating an accumulated weight by accumulating the weight assigned to each of the plurality of attributes; and

(d) ranking the plurality of text items based on the accumulated weight of each of the plurality of texts.

7. The system of claim 6, wherein the plurality of attributes includes at least one of style, font size, and location of the text item.

8. The system of claim 6 further forming a plurality of search terms based on a result of the ranking step.

9. A machine-readable medium bearing instructions for obtaining representative text items from a plurality of text items in an active task, each of the plurality of text items having a plurality of attributes, the instructions upon execution by a data processing system causing the data processing system to perform the steps of:

(a) for each of the plurality of text items, identifying each of the plurality of attributes;

(b) for each of the plurality of text items, assigning a weight to each of the plurality of attributes, wherein at least two of the plurality of attributes are assigned different weights;

(c) for each of the plurality of text items, calculating an accumulated weight by accumulating the weight assigned to each of the plurality of attributes; and

(d) ranking the plurality of text items based on the accumulated weight of each of the plurality of texts.

10. A method for obtaining representative text items from a plurality of text items in an active task, comprising the steps of:

- (a) identifying attributes for each of the plurality of text items; and
- (b) excluding text items having at least one of the attributes consisting of:
 - containing less than n letters or characters, wherein n is a tunable number, unless the text items are part of an exception list or part of recognized constituent items;
 - containing all numbers, unless the text items are part of an exception list or part of recognized constituent items;
 - part of a stop list;
 - part of a stop list including text items corresponding to a specific user;
 - part of a stop list including text items corresponding to an information source; and
 - part of a link to retrieve a file or a web page.

11. A machine-readable medium bearing instructions for obtaining representative text items from a plurality of text items in an active task, the instructions upon execution by a data processing system causing the data processing system to perform the steps of:

- (a) identifying attributes for each of the plurality of text items; and
- (b) excluding text items having at least one of the attributes consisting of:

containing less than n letters or characters, wherein n is a tunable number, unless the text items are part of an exception list or part of recognized constituent items;

containing all numbers, unless the text items are part of an exception list or part of recognized constituent items;

part of a stop list;

part of a stop list including text items corresponding to a specific user;

part of a stop list including text items corresponding to an information source; and

part of a link to retrieve a file or a web page.

12. A method for retrieving information related to the context of a plurality of text items, wherein each of the plurality of text items has a plurality of attributes, the method comprising the steps of:

(a) for each of the plurality of text items, identifying each of the plurality of attributes;

(b) for each of the plurality of text items, assigning a weight to each of the plurality of attributes, wherein at least two of the plurality of attributes are assigned different weights;

(c) for each of the plurality of text items, calculating an accumulated weight by accumulating the weight assigned to each of the plurality of attributes;

(d) ranking the plurality of text items based on the accumulated weight of each of the plurality of text items;

(e) generating a set of search terms containing a predetermined number of text items based on rankings of the plurality of text items; and

(f) initiating an information retrieval process based on the set of search terms.

13. A machine-readable medium bearing instructions for retrieving files that are related to the context of a plurality of text items, wherein each of the plurality of text items has a plurality of attributes, the instructions upon execution by a data processing system controlling the data processing system to perform the steps of:

(a) for each of the plurality of text items, identifying each of the plurality of attributes;

(b) for each of the plurality of text items, assigning a weight to each of the plurality of attributes, wherein at least two of the plurality of attributes are assigned different weights;

(c) for each of the plurality of text items, calculating an accumulated weight by accumulating the weight assigned to each of the plurality of attributes;

(d) ranking the plurality of text items based on the accumulated weight of each of the plurality of text items;

(e) generating a set of search terms containing a predetermined number of text items based on rankings of the plurality of text items; and

(f) initiating an information retrieval process based on the set of search terms.

14. A method for formulating context representations comprising the steps of:
identifying a relationship between a first text item and a second text item in a first text item group;
converting the first text item group into a second text item group by applying at least one conversion rule;
determining whether the first text item is in the second text item group;
in response to the first text item being in the second text item group, accessing information related to the relationship between the first text item and the second text item;
determining whether the second text item is in the second text item group;
in response to the second text item being in the second text item group, grouping the first text item and second text item.

15. The method of claim 14 further indexing the relationship between the first text item and the second text item using the first text item.

16. The method of claim 14, wherein the grouping step groups the first text item and the second text item based on the relationship between the first text item and the second text item.

17. The method of claim 16, wherein the first text item and the second text item appear next to each other in the first text item group.

18. The method of claim 17, wherein the first text item and the second text item are group next to each other to form a single search term.

19. The method of claim 14 further conducting an information search process using the grouped first text item and second text item.

20. A data processing system for formulating context representations comprising:

- a data processor for processing data;

- a data storage device for storing instructions; and

- a data transmission path coupled to the data processor and the data storage device;

wherein the instructions, when executed by the data processor, controls the data processing system to perform the machine-implemented steps of:

- identifying a relationship between a first text item and a second text item in a first text item group;

- converting the first text item group into a second text item group by applying at least one conversion rule;

- determining whether the first text item is in the second text item group; in response to the first text item being in the second text item group, accessing information related to the relationship between the first text item and the second text item;

- determining whether the second text item is in the second text item group;

in response to the second text item being in the second text item group,
grouping the first text item and second text item.

21. A machine-readable medium bearing instructions for formulating context representations, the instructions upon execution by a data processing system causing the data processing system to perform the steps of:

identifying a relationship between a first text item and a second text item in a first text item group;

converting the first text item group into a second text item group by applying at least one conversion rule;

determining whether the first text item is in the second text item group;

in response to the first text item being in the second text item group, accessing information related to the relationship between the first text item and the second text item;

determining whether the second text item is in the second text item group;

in response to the second text item being in the second text item group, grouping the first text item and second text item.

22. A method for determining an emphasis status of a text item comprising the steps of:

accessing information related to an attribute value of each of a plurality of text items;

generating a representative attribute value of the plurality of text items based on the attribute value of each of the plurality of text item;
accessing information related to the attribute value of the text item;
comparing the attribute value of the text item with the representative attribute value of the plurality of text items; and
determining the emphasis status of the text item based on a result of the comparing step.

23. The method of claim 22, wherein the attribute value is selected from the group consisting of font size and line height.

24. The method of claim 22, wherein the text item is considered as emphasized in response to the attribute value of the text item being larger than the representative attribute value.

25. An adaptive method for generating text properties based on a plurality of text items in an active task, each of the plurality of text items having a plurality of attributes, the method comprising the steps of:

- (a) determining properties of the active task;
- (b) for each of the plurality of text items, identifying each of the plurality of attributes;

- (c) for each of the plurality of text items, assigning a weight to each of the plurality of attributes, wherein at least two of the plurality of attributes are assigned different weights, and the weight is tunable based on the properties of the active task;
- (d) for each of the plurality of text items, calculating an accumulated weight by accumulating the weight assigned to each of the plurality of attributes; and
- (e) ranking the plurality of text items based on the accumulated weight of each of the plurality of text items.

26. The method of claim 25, wherein the properties of the active task include at least one of application software being employed to perform the active task, the type or genre of the active task, attributes related to the user manipulating the active task, properties of an information source on which a search will be conducted, and the state of the active task.

27. A method for obtaining representative text items based on a plurality of text items of an active task, comprising the steps of:

- (a) determining properties of the active task;
- (b) identifying attributes for each of the plurality of text items; and
- (c) based on the properties of the active task, excluding text items by applying at least one exclusion rule.

28. The method of claim 27, wherein the at least one exclusion rule excludes text items having at least one of the attributes consisting of:

containing less than n letters or characters, wherein n is a tunable number, unless the text items are part of an exception list or part of recognized constituent items;

containing all numbers, unless the text items are part of an exception list or part of recognized constituent items;

part of a stop list;

part of a stop list including text items corresponding to a specific user;

part of a stop list including text items corresponding to an information source; and

part of a link to retrieve a file or a web page.

29. The method of claim 27, wherein the properties of the active task include at least one of application software being employed to perform the active task, the type or genre of the active task, attributes related to the user manipulating the active task, properties of an information source on which a search will be conducted, and the state of the active task.